

# Het effect van een interventie nagaan. Een illustratieve casus met toepassing van tijdreeksenanalyse en randomiseringstoetsen

S. Joos<sup>1</sup>, P. Onghena<sup>2</sup> en G. Pieters<sup>1</sup>

## Inleiding

De vraag of therapie effectief is heeft al heel wat pennen in beweging gebracht, wat gezien de complexiteit van de vraag niet verwonderlijk is. Een overzicht van de pogingen om deze vraag te beantwoorden leert ons dat een sluitend antwoord voorlopig ontbreekt maar dat de vraag zich steeds preciezer stelt.

Eysenck (1952) stelde de vraag zeer algemeen. Is psychotherapie effectief ja of nee? Hij stelde dat ongeveer twee derde van de patiënten met psychoneurotische problemen spontaan herstelt, ook zonder behandeling. De gerapporteerde verbeteringspercentages na psychotherapie overstijgen dit getal niet. Deze krasse stellingname lokte polemiek uit maar betekende ook de start van de kritische evaluatie van psychotherapie.

Daarna volgde de fase van het vergelijkend therapie-onderzoek. Is gedragstherapie effectiever dan psychoanalytische therapie? Het werd de aanzet tot een bittere scholenstrijd. Ieder probeerde zijn gelijk te bepleiten en de argumentatie van de tegenpartij te ontzenuwen. Het werd een strijd zonder winnaar. 'Everyone has won and all must have prizes!' De grote verschillen in conceptualisatie van de problematiek tussen de diverse therapeutische scholen vertalen zich niet in grote verschillen in effectiviteit. De gelijke resultaten van verschillende therapeutische scholen leidden tot het beklemtonen van de 'aspecifieke' factoren als therapeutisch werkzame factoren. Tenslotte doen we allemaal dezelfde dingen. Met dit gelijkspel zijn sommigen meer tevreden dan anderen (Sloane, Staples, Cristol et al., 1975; Luborsky, Singer & Luborsky, 1975; Smith, Glass & Miller, 1980).

1 Universitair Centrum St.-Jozef, Leuvensesteenweg 517, B-3070 Kortenberg.

2 Faculteit Psychologie en Pedagogische Wetenschappen, KU Leuven.



Intussen gaat de zoektocht naar specifieke therapeutisch werkzame factoren onverminderd voort. Dit gaat gepaard met een evolutie van aandacht voor het algemene naar het bijzondere en individuele. De vraagstelling is steeds meer gedifferentieerd: welke therapeut behaalt een bepaald resultaat bij welke patiënt met welke problematiek door middel van welke interventie? De therapeutische relatie wordt een trefpunt van talloze interagerende therapeut- en cliëntvariabelen. Elke therapie is complex en volstrekt uniek. De gevalsstudie wordt weer populair met respect voor de talloze individuele variabelen. Maar is er nog wetenschap als we bij het strict individuele belanden? De zoektocht is niet ten einde maar de plaats waar gezocht wordt is steeds meer de concrete therapeutische praktijk.

Dit betekent dat vooral de therapeut in de juiste positie is om iets te zeggen over het effect van een therapie. Dit veronderstelt dat hij naast een therapeutische belangstelling ook interesse heeft voor het effect van wat hij doet. Idealiter is hij niet alleen therapeut maar ook kritisch onderzoeker. Zeker de gedragstherapeut voelt zich goed in de rol van 'scientist-practitioner'. Vanuit dit in Amerika populaire model stelt men dat een therapeut niet alleen de nodige therapeutische vaardigheden moet hebben maar ook de nodige methodologische kennis om het effect van een interventie op een verantwoorde manier na te gaan. Men spreekt van 'N=1'-onderzoek, 'single subject'-studie of 'single-case experiment' (Hilliard, 1993; McLeod, 1994).

We stellen hierna een gevalsstudie voor, waarbij we niet zozeer stilstaan bij de methodologische problemen (interne en externe validiteit) die de studie van één geval stelt (Kazdin, 1982), maar eerder bij de mogelijkheden wat betreft data-analyse. We introduceren de mogelijkheden van statistische methoden (tijdreeksenanalyse en randomiseringstoetsen) die flexibel toepasbaar zijn in het individuele geval en gemakkelijk in de therapeutische praktijk te implementeren zijn.

Onze invalshoek is pragmatisch van aard en individu-gericht. Op basis van de probleemanalyse en de beschikbare onderzoeksgegevens wordt een therapeutische strategie gekozen. De werkzaamheid van de geplande therapie in het individuele geval kan worden getoetst. Zijn er aanduidingen dat onze therapeutische interventie effectief is? Zo beschouwd kan elke therapie worden gezien als een uniek experiment, waarbij de therapeutische interventie de kritische toets van onze werkhypothese is.



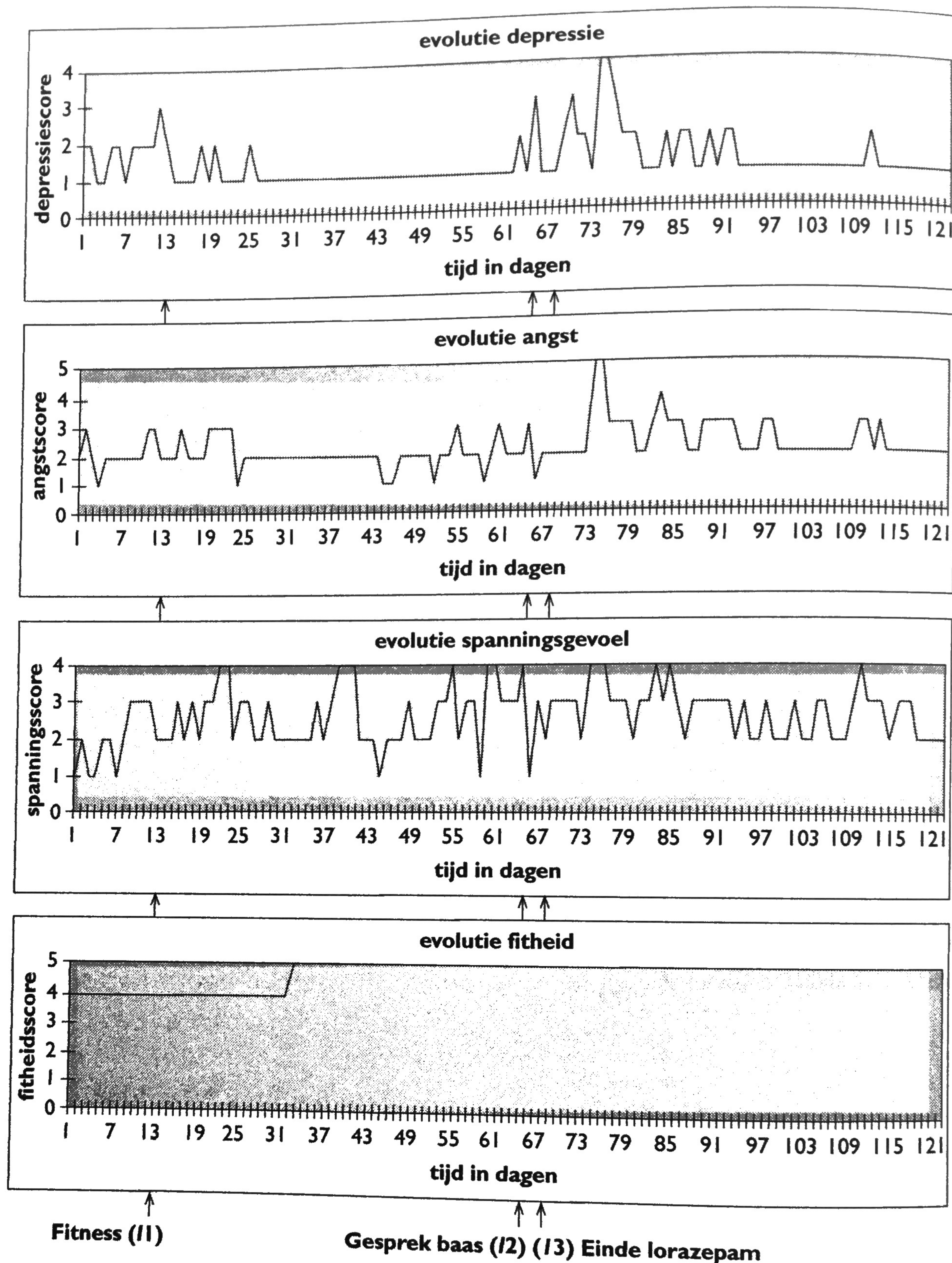
Alex is een 45-jarige technicus, die reeds herhaaldelijk wegens depressie werd opgenomen in een psychiatrische kliniek. Momenteel verblijft hij op een afdeling voor gedragstherapie. De DSM-IV-diagnose op het ogenblik van de studie vermeldt op As I als klinische symptomen matige depressie in engere zin met recidiverend karakter, diffuse angststoornis en misbruik van benzodiazepinen. Op As II wordt een persoonlijkheidsstoornis met afhankelijke en theatrale kenmerken vermeld. Alex is meerbegaafd (WAIS-IQ 122).

Onze gevalsstudie start op het ogenblik dat het gedragstherapeutisch team de beslissing heeft genomen om de depressieve klachten van Alex te behandelen met fitnessstraining. Er zijn aanduidingen in de literatuur dat fitnessstraining effectief is als behandeling bij depressie (Martinsen, 1995; Van Coppenolle, Pieters, Knapen & Peuskens, 1993). In ons onderzoek bestuderen we de samenhang tussen depressie en fitnessstraining. Preciezer geformuleerd is de vraag: hoe effectief is fitnessstraining in het reduceren van depressieve gevoelens? Er wordt vooral aandacht besteed aan de methodische mogelijkheden die ter beschikking zijn om een dergelijke vraagstelling in het individuele geval op een verantwoorde en boeiende manier te benaderen. We toetsen de voorspelling dat fitnessstraining de depressieve gevoelens van Alex zal verminderen. Als de gemiddelde depressieve patiënt goed reageert op een fitnessprogramma, is dat dan ook het geval bij Alex?

De verzamelde gegevens vinden we in figuur 1. Vier klachten werden door Alex dagelijks geregistreerd: depressie, angst, spanning en fitheid. In grafische vorm zien we de evolutie van deze vier variabelen over een periode van 123 dagen. De depressiescore, die ons vooral interesseert, geeft het subjectieve gevoel van depressie bij Alex weer. Er werd hem gevraagd om dagelijks op een 7-puntsschaal de ernst van zijn depressie aan te geven (1 = helemaal niet, 2 = weinig, 3 = in lichte mate, 4 = tamelijk, 5 = duidelijk, 6 = erg, 7 = uitermate depressief). Dezelfde subjectieve schaal werd gebruikt om de evolutie van angst, spanning en fitheid aan te duiden. Het voordeel is dat een dergelijke gegevensverzameling zeer eenvoudig is en inzicht geeft in ernst en evolutie van de probleemvariabelen.

De pijlen in figuur 1 markeren belangrijke gebeurtenissen, waarvan een effect op de betrokken variabelen wordt verwacht. De eerste interventie, I I in figuur 1, vindt plaats op dag 15, wanneer de fitnessstraining start. Deze training wordt gegeven door de psy-





Figuur 1. Voorstelling van de verzamelde gegevens.

chomotorisch therapeut. Het oefenprogramma loopt over een periode van 10 weken met per week 3 sessies. Er wordt geoefend op een looptapijt, een ergometerfiets en een krachtaapparaat. Tevens worden ontspanningsoefeningen en stretching ingelast. In de aanvangsfase is de belasting laag. De oefeningen worden geleidelijk opgevoerd, maar blijven onder de vermoeidheidsdrempel. Voor dat vermoeidheid optreedt wordt een rustpauze ingelast. De opvolging inspanning-rust komt in de plaats van de opeenvolging activiteit-vermoeidheid-rust. Het is de bedoeling dat de patiënt evolueert van een vermoeidheids- naar een activiteitsoriëntatie (Knapen, 1995).



De tweede pijl, op dag 61, geeft een belangrijk gesprek met de baas van Alex aan over eventuele werkhervatting. Na enkele mislukte pogingen om het werk te hervatten is dit een moment van confrontatie met de verwachtingen van de externe realiteit (Interventie 2 of I 2 in figuur 1).

De derde pijl, op dag 64, geeft aan wanneer met lorazepam werd gestopt. Gezien de jarenlange afhankelijkheid van benzodiazepinen is dit een belangrijk moment. Vanaf nu staat Alex er in zekere zin 'alleen' voor (Interventie 3 of I 3 in figuur 1).

Kunnen we op basis van de verzamelde gegevens iets zeggen over het effect van deze drie belangrijke gebeurtenissen (I 1, I 2, I 3) op het verloop van de vier klachten van Alex? In principe is dit mogelijk. We beperken ons vooral tot bestudering van de invloed van I 1, de fitnessstraining, op het verloop van de depressie.

## **Data-analyse**

### **1. Visuele inspectie**

Wat leert ons een visuele inspectie van het verloop van de depressiescores in figuur 1? Na I 1, het invoeren van de fitnessstraining, dalen de depressieve gevoelens. Bij I 2 en I 3 nemen ze weer toe om vervolgens weer te dalen. Kennelijk reduceert fitnessstraining de depressieve gevoelens maar I 2 en I 3 doen deze daling teniet. Na een tijdelijke opflakking van ongeveer een maand dalen de depressiescores weer. Is dan het verstorende effect van I 2 en I 3 uitgewerkt? De angstgevoelens evolueren ongeveer op dezelfde manier. De spanningsgevoelens hebben eerder de neiging om na I 1 stelselmatig toe te nemen. De fitheid evolueert na een aantal weken fitnessstraining snel naar een stabiel niveau.

Deze globale aanduidingen zijn nuttig maar niet precies genoeg. Om te kunnen zeggen of de daling van de depressiescores na het invoeren van de fitnessstraining ook statistisch significant is, of verschillend van een toevallige fluctuatie moet er statistisch worden getoetst (Park, Marascuilo & Gaylord, 1990).

### **2. Tijdreeksenanalyse**

#### *Korte situering*

Een tijdreeks bestaat uit een opeenvolgende reeks waarnemingen van een variabele op verschillende tijdstippen. De dagelijkse depressiescores van Alex bijvoorbeeld vormen een tijdreeks. Hoe



evolueert de depressie in functie van de tijd? Bij tijdreeksenanalyse gaat men kijken of er een bepaald patroon in de tijdreeks aanwezig is. Welke factoren bepalen het verloop van bijvoorbeeld de opeenvolgende reeks depressiescores? De fitnessstraining is één van de factoren waarvan we een effect op het verloop van de depressiescores verwachten. Wellicht spelen ook de andere interventies een rol. Om een verantwoorde uitspraak te kunnen doen over het effect van een interventie moet in eerste instantie een diagnose van de tijdreeks worden gemaakt. Het onderliggende patroon moet allereerst worden verduidelijkt. Dit veronderstelt een voldoende aantal observaties, minimaal 50 à 100, wat in de praktijk niet altijd haalbaar is. In principe zijn continue variabelen noodzakelijk. De depressiescores zijn wel niet echt continu maar ze worden gewoonlijk wel zo behandeld.

Een typerend gegeven voor tijdreeksen is dat de opeenvolgende waarnemingen onderling niet onafhankelijk behoeven te zijn. Het lijkt aannemelijk dat de opeenvolgende depressiescores een zekere samenhang vertonen met elkaar. Als Alex zich gisteren erg depressief voelde, dan verwachten we dat de hoge depressiescore van gisteren ook de depressiescore van vandaag mede zal bepalen. En misschien vertonen ook verder in de tijd liggende depressiescores een samenhang met de huidige depressiescore. Om deze onderlinge afhankelijkheid te bestuderen is de *autocorrelatiecoëfficiënt* een belangrijk hulpmiddel. Een autocorrelatiecoëfficiënt met stapwaarde 1 betekent dat de correlatie wordt berekend tussen de waargenomen tijdreeks en dezelfde tijdreeks, die één tijdstip verschuift. Verschuivingen over 2, 3, 4 of meer ( $k$ ) tijdstippen levert autocorrelatiecoëfficiënten in respectievelijk stap 2, stap 3, stap 4 en stap  $k$ . Normaal daalt de autocorrelatie naarmate het aantal stappen toeneemt. De grafische voorstelling van de verschillende autocorrelatiecoëfficiënten van een tijdreeks noemen we een autocorrelogram (Buijs, 1994). Naast de autocorrelatiecoëfficiënt kan ook de *partiële autocorrelatiecoëfficiënt* berekend worden vanaf een stapwaarde van 2 (verschuiving van de observatiebeurten over 2 tijdstippen). Hierbij wordt in de autocorrelatie het aandeel van de tussenliggende observaties uitgezuiverd. Op basis van beide autocorrelatiecoëfficiënten wordt de diagnose van een tijdreeks mogelijk. Dit wil zeggen dat men de tijdreeks poogt te scheiden in een aantal componenten.

#### A. Trend

Eerst en vooral gaat men na of er een trend in de tijdreeks aanwezig is. Is er in de gegevens een algemene tendens tot stijgen of dalen merkbaar? We noemen een tijdreeks stabiel of stationair wan-



neer het niveau en de spreiding in de loop van de tijd niet veranderen. Trend is een vorm van instabiliteit (I) en betekent dat er in de gegevens een stelselmatige verandering aanwezig is. Toevallige veranderingen zijn aanwijzingen dat er sprake is van een proces op drift. Als de trendlijn bepaald is, bijvoorbeeld door het opstellen van een regressievergelijking, dan kunnen systematische afwijkingen van deze trendlijn worden bestudeerd. We denken onder meer aan periodiek terugkerende verhogingen of verlagingen in de tijdreeks, bijvoorbeeld wekelijks, maandelijks of seizoengebonden.

Een eerste 'behandeling' van de tijdreeks bestaat erin ze stabiel of stationair te maken. Wanneer het niveau van de tijdreeks verandert kan differentiatie nodig zijn. Differentiëren betekent dat men niet met de oorspronkelijke tijdreeks werkt maar wel met de verschillscores, die men verkrijgt door van elke score in de tijdreeks de vorige score af te trekken. De nieuwe gedifferentieerde tijdreeks geeft dan geen beeld van de depressiescore per dag, maar van het verschil in depressiescore tussen twee dagen. Wanneer de spreiding van de tijdreeks niet stabiel blijft kan een correctie nodig zijn door de oorspronkelijke scores logaritmisch te transformeren.

## B. Afhankelijkheid

Met De Boeck (1988) onderscheiden we twee vormen van afhankelijkheid: nawerking en overheveling. We spreken van *nawerking* wanneer de waarde van een tijdreeksvariabele samenhang vertoont met de waarde van die tijdreeksvariabele op eerdere tijdstippen. In statistische termen spreekt men van autoregressie (AR). *Overheveling* betekent dat invloeden die een vorige score bepalen ook de volgende score medebepalen. In statistische termen spreekt men van het zich verplaatsend gemiddelde ('moving average', MA).

Op basis van de verzamelde waarden in de tijdreeks wordt een zogenoemd ARIMA-model opgesteld. Men noemt dit de Box-Jenkins-methode. De onderzoeker bepaalt de waarde van de nawerkingsfactor (p), de waarde van de overhevelingsfactor (q) en de differentiatiewaarde (d). Deze d-waarde geeft aan hoe vaak men de tijdreeks moet differentiëren om er een stationaire tijdreeks van te maken. Voor een introductie in de diagnosestelling en de behandeling van een tijdreeks verwijzen we naar McCleary en Hay (1980) en de gespecialiseerde handleiding bij het gebruikte computerprogramma (SAS/ETS user's guide).



### C. Het effect van een interventie meten

Wanneer de tijdreeks door een adequaat ARIMA-model behandeld is mogen de overblijvende scores geen autocorrelatie meer vertonen. Door middel van de Durbin-Watson-toets kan worden nagegaan of er nog sprake is van autocorrelatie. Daarna kan de bijdrage van een interventie aan de tijdreeks worden geschat. De aard van het interventie-effect kan in een interventiemodel gespecificeerd worden. Verwacht men op het ogenblik van de interventie een onmiddellijk effect dat blijft voortduren? Of verwacht men een direct effect, dat langzaam uitsterft? Afhankelijk van intreden en duur van het effect onderscheidt De Boeck (1986) vijf interventiemodellen: geleidelijk en permanent effect, abrupt en permanent effect, geleidelijk en tijdelijk effect, abrupt en tijdelijk effect, en ten slotte ogenblikkelijk verdwijnend effect.

### D. Het effect van een parallel verlopende tijdreeks

Het is ook mogelijk om na te gaan of een tijdreeks wordt beïnvloed door een andere, gelijktijdige tijdreeks. Men spreekt hier van transfer van de ene naar de andere tijdreeks. Bijvoorbeeld: wat is bij Alex de weerslag van het spanningsgevoel of het gevoel van fitheid op de depressie? Men spreekt dan van multivariate tijdreeksenanalyse. De relatie tussen twee of meer tijdreeksen kan in een model worden gegoten. Hoe beïnvloedt de tijdreeks spanning de tijdreeks depressie bij Alex? Om de afhankelijkheid binnen een tijdreeks te bepalen steunden we op de autocorrelatiecoëfficiënt. Om de afhankelijkheid tussen tijdreeksen te identificeren steunen we op de cross-correlatiefunctie. Vooraf moet een adequaat ARIMA-model voor beide tijdreeksen worden opgesteld, gezien er ook in de input-tijdreeks (spanning) sprake kan zijn van onstabielheid en afhankelijkheid. Daarna kan de cross-correlatie worden berekend en dit voor verschillende stapwaarden, waarbij de beide tijdreeksen zowel voorwaarts als achterwaarts ten opzichte van elkaar worden verschoven. Dit laat toe om niet alleen iets te zeggen over de sterkte van het verband tussen de twee tijdreeksen maar ook over de richting. Beïnvloedt het verloop van de depressie het spanningsverloop of heeft de spanning een weerslag op de depressie?

### *Praktische toepassing*

Het analyseren van een tijdreeks veronderstelt een specifiek computerprogramma. Hier hebben we de aparte ETS-module van SAS 6.08 voor Windows gebruikt, via de licentie van het Universitair Rekencentrum van de Katholieke Universiteit Leuven.



Tabel 1. Autocorrelaties en partiële autocorrelaties van de variabele depressie.

## Autocorrelaties

'lag'	covariantie	correlatie	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
0	0,382048	1,00000																					
1	0,180580	0,47266									.												
2	0,133583	0,34965									.												
3	0,100203	0,26228								.													
4	0,123734	0,32387								.													
5	0,117388	0,30726								.													
6	0,086652	0,22681								.													
7	0,085792	0,22456								.													
8	0,038795	0,10155								.													
9	0,073100	0,19134								.													
10	0,050494	0,13217								.													
11	0,060408	0,15812								.													
12	0,032316	0,08459								.													
13	0,028614	0,07490								.													
14	0,019624	0,05136								.													
15	0,010634	0,02783								.													
16	0,0016439	0,00430								.													
17	-0,0073460	-0,01923								.													
18	-0,029952	-0,07840								.		**											
19	-0,038942	-0,10193								.		**											
20	-0,020898	-0,05470								.		*											
21	-0,046148	-0,12079								.		**											
22	-0,047008	-0,12304								.		**											
23	-0,064128	-0,16785								.		**											
24	-0,064987	-0,17010								.		**											

## Partiële autocorrelaties

'lag'	correlatie	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
1	0,47266									.												
2	0,16256									.												
3	0,06222									.												
4	0,19262									.												
5	0,09679									.												
6	-0,01837									.												
7	0,06500									.												
8	-0,12073									.	**											
9	0,11033									.												
10	-0,02693									.	*											
11	0,03669									.												



'lag'	correlatie	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9
12	-0,03356									.	*			.							
13	-0,00863									.				.							
14	-0,03674									.	*			.							
15	-0,02060									.				.							
16	-0,06474									.	*			.							
17	-0,00029									.				.							
18	-0,11729									.	**			.							
19	-0,03013									.	*			.							
20	0,02725									.		*		.							
21	-0,09201									.				.							
23	-0,05139									.	*			.							
24	-0,06420									.	*			.							

De puntlijn markeert het betrouwbaarheidsinterval.

### A. Trend

In tabel I vinden we de gegevens die ons toelaten een diagnose te stellen van de tijdreeks van de depressiescores van Alex. De computer geeft standaard 24 autocorrelatie- en partiële autocorrelatiecoëfficiënten. In stapwaarde 1 (=lag 1) is de autocorrelatie 0,47. Met het progressief verhogen van de stapwaarde tot 24 zien we de autocorrelaties geleidelijk dalen en vanaf stapwaarde 17 zelfs negatief worden. Deze evolutie wordt in beeld gebracht door de sterretjes. De verkregen coëfficiënten zijn niet zo hoog en tonen geen typerende convexe daling met het toenemen van de stapwaarde. We beschouwen deze reeks derhalve als stationair en er is dus geen differentiatie nodig om een geïntegreerd proces in de tijdreeks te herleiden tot een stationair proces (door met de verschilcores te werken).

### B. Afhankelijkheid

Typerend in deze tijdreeks is dat de autocorrelatiecoëfficiënten dalen van 0,47 in stapwaarde 1 en 0,35 in stapwaarde 2 tot 0,26 in stapwaarde 3 om vervolgens weer te stijgen naar 0,32 in stapwaarde 4. Daarna volgt opnieuw een daling en verschillen de verkregen coëfficiënten niet langer van 0. De knik bij een stapwaarde van 4 wijst op een nawerkingsfactor. Dit betekent dat de depressiescore van Alex niet alleen bepaald wordt door de depressiescore van de vorige dag, maar ook door de score van vier dagen geleden. We zien dit bevestigd in de partiële autocorrelatiecoëfficiënten, die statistisch significant zijn bij stapwaarde 1 en 4. Deze tijdreeks wordt het best weergegeven door een autoregressief model. De geschatte waarden in het ARIMA-model zijn voor AR (1, 4), voor I (0) en voor MA (0). Toepassing van dit model op de tijdreeks her-



leidt de residuele reeks tot witte ruis of een reeks zonder autocorrelatie. Het mathematisch model voor de tijdreeks depressie, met de geschatte autoregressieve parameters, ziet er als volgt uit:

$$\text{Depressie (t)} = 1,34 + 0,42 \text{ depressie (t-1)} + 0,22 \text{ depressie (t-4)} + \epsilon (t)$$

Depressie (t) is de depressiescore op een bepaald tijdstip. Depressie (t-1) is de depressiescore van de dag voordien. Depressie (t-4) is de depressiescore van 4 dagen voordien. De waarde 1,34 is een gemiddelde term. De waarde 0,42 is de geschatte autoregressieve parameter, het gewicht dat de score van de dag voordien krijgt. De waarde 0,22 is het gewicht van de depressiescore van vier dagen geleden. In de vergelijking weegt deze waarde duidelijk minder door dan die van de onmiddellijk voorafgaande score.  $\epsilon (t)$  is de 'random' foutenterm.

### C. Het effect van fitnessstraining

We zijn nu klaar om een interventiemodel op te stellen. De fitnessstraining startte op dag 15 in de tijdreeks van de depressiescores. We opteren voor een eenvoudig model, waarbij de fitnessstraining vanaf dag 15 de tijdreeks van de depressie beïnvloedt. Hiertoe wordt er een parallel verlopende tijdreeks 'fitnessstraining' geconstrueerd, die voor dag 15 de waarde 0 aanneemt en vanaf dag 15 de waarde 1. We kunnen nu nagaan wat, naast het ARIMA-model, de bijdrage is van dit interventiemodel aan de tijdreeks depressie. Het beste model wordt weergegeven door de volgende formule:

$$\text{Depressie (t)} = 1,73 + 0,39 \text{ depressie (t-1)} + 0,21 \text{ depressie (t-4)} - 0,46 \text{ fitheid (t)}$$

We zien dat de fitheidsscore op tijdstip t bijdraagt aan de reductie (negatief gewicht) van de depressiescore op tijdstip t, al is de verkregen T-ratio niet statistisch significant (T-ratio: -1,68). De positieve effecten van de fitnessstraining worden hier wellicht verdoezeld door het negatieve effect van I2 en I3 op het verloop van de depressie.

### D. Het effect van het spanningsgevoel op depressie

Na de voorbehandeling van depressie en spanning voor afhankelijkheid en instabiliteit, 'prewhitening' genoemd, blijkt dat de cross-correlatiecoëfficiënt slechts in stapwaarde 0 (zonder verschuiving dus) significant is (0,36). Er is een directe samenhang tus-



sen spanning en depressie, waarbij het in dit geval onmogelijk is de richting van het verband aan te geven. Een goed model voor de tijdreeks depressie is:

$$\text{Depressie}(t) = 1,11 + 0,39 \text{ dep}(t-1) + 0,17 \text{ dep}(t-4) + 0,32 \text{ spanning}(t) - 0,69 \text{ fitheid}(t)$$

Al de betrokken variabelen in dit model hebben een significante bijdrage. De respectievelijke T-ratio's zijn 4,36, 4,57, 2,02, 5,47 en -2,83. Als Alex zich gespannen voelt dan draagt dit bij tot zijn depressiescore, fitnessstraining vermindert de depressiescore. De analyse van de cross-correlatiecoëfficiënten tussen fitness en depressie leert ons dat de waarden van stapwaarde 0 (-0,27) geleidelijk dalen (stapwaarde 1, -0,23), (stapwaarde 2, -0,19), (stapwaarde 3, -0,19) enzovoort om in stapwaarde 8 niet langer van 0 te verschillen. Dit wijst op een positief effect van fitnessstraining op depressie dat over een periode van ongeveer een week langzaam afneemt. Het stoppen met lorazepam en het gesprek met de baas leveren een te verwaarlozen bijdrage aan het depressieverloop. De significante cross-correlatiecoëfficiënt in stapwaarde +2 (0,18) geeft een aanduiding dat het gesprek met de baas na 2 dagen gevolgd wordt door een verhoging van de depressiescore.

### 3. Randomiseringstoets

#### *Korte situering*

Randomiseringstoetsen zijn nuttig om het effect van een interventie te meten. Het zijn vooral Edgington (1984) en in ons taalgebied De Boeck (1988) en Onghena (1992) die gewezen hebben op de interessante toepassingen bij N=1-onderzoek. Randomiseringstoetsen zijn niet-parametrisch en geven geen veronderstellingen over normaliteit, homogeniteit van varianties en onafhankelijkheid. De enige veronderstelling die gedaan wordt, is dat een therapie gepland wordt volgens een randomiseringsmodel, dat wil zeggen dat het tijdstip van de interventie louter van het toeval afhangt. We zijn geïnteresseerd in het effect van de fitnessstraining op de depressie. Dit impliceert dat het tijdstip waarop de fitnessstraining start van niets anders afhangt dan van het toeval. Er is een basislijnconditie (A) en een conditie na de interventie (B). We beperken ons hier tot een A-B-design, al zijn allerlei soorten designs mogelijk.

Beschouwen we de 123 depressiescores van Alex. De toevallige start van de fitnessstraining op dag 15 creëert een A- en een B-fase.



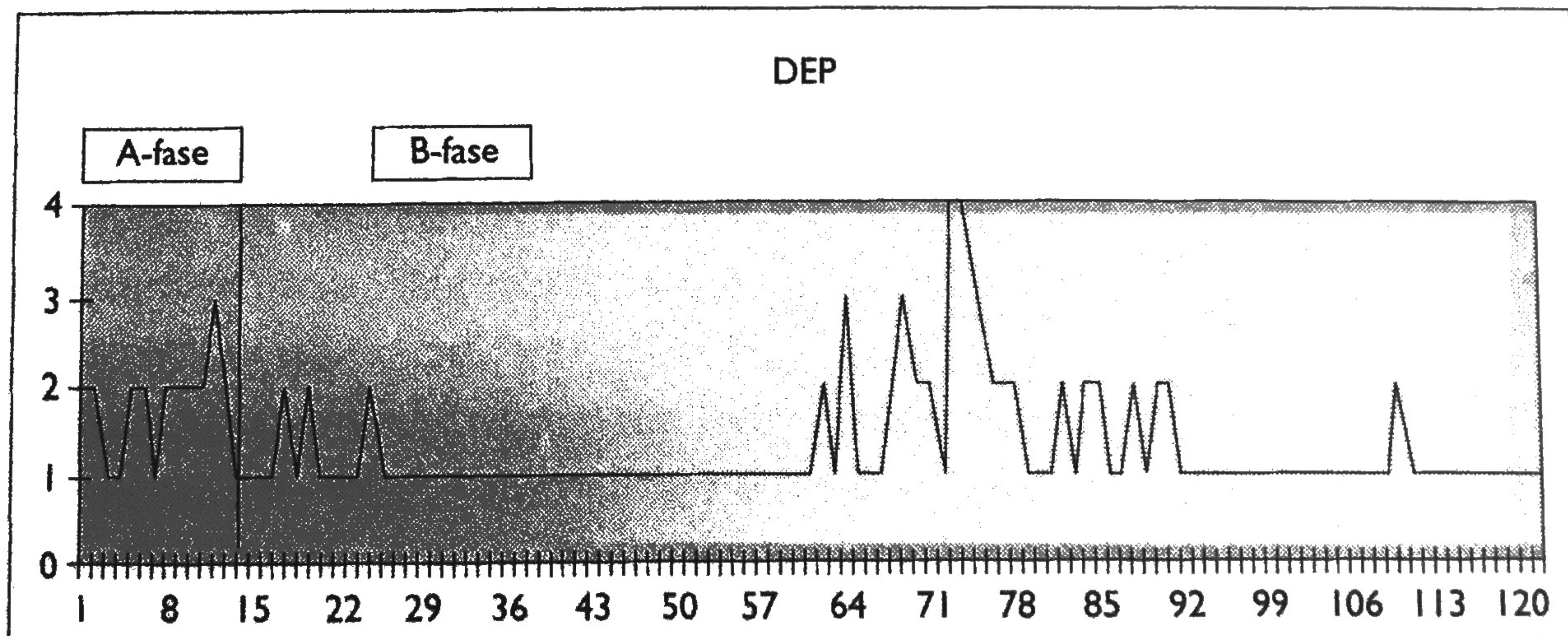
Het spreekt vanzelf dat een minimumaantal observaties per fase noodzakelijk is. In dit geval kozen we voor een minimum van 10 observaties per fase. Rekening houdend met deze marge komt, volgens het randomiseringsmodel, elke dag in aanmerking om de fitnessstraining te starten, na 10 dagen, na 11 dagen, enzovoort, met als laatste mogelijkheid na 113 dagen. De verschillende mogelijke verdelingen zijn: 10A/113B, 11A/112B, 12A/111B, .... 112A/11B, 113A/10B. Op grond van het toeval selecteerden we bij Alex 14A/109B.

Als statistiek kiezen we voor het verschil tussen de gemiddelden van fase A en fase B. Om het statistisch toetsen mogelijk te maken is het nodig dat we de verdeling van deze statistiek kennen onder de nulhypothese, in casu dat de interventie geen effect heeft. Deze verdeling wordt per geval op basis van het randomiseringsmodel en op basis van de verzamelde gegevens berekend. Allereerst worden al de mogelijke verschillen tussen de gemiddelden van fase A en fase B (10A/113B, 11A/112B .... 112A/11B, 113A/10B) berekend. Daarna worden deze verschillen naar grootte geordend. Vervolgens wordt een cumulatieve probabiliteitsdistributie opgesteld. In deze distributie kan worden afgelezen wat de probabilmiteit is om een bepaald verschil tussen fase A en fase B te vinden. Belangrijk is dat per geval een verdeling op maat wordt gemaakt op basis van de verzamelde gegevens.

### *Praktische toepassing*

Allereerst wijzen we erop dat het uitvoeren van randomiserings-toetsen meestal de beschikbaarheid van een computer en aangepaste programmatuur veronderstelt. Het aantal randomisaties of permutaties kan hoog oplopen, zeker wanneer complexere designs worden gebruikt (zoals ABA en ABAB). In deze toepassing maakten we gebruik van het computerprogramma Single Case Randomization Tests, versie 1.1 (Van Damme & Onghena, 1993).

Wat leert ons het toepassen van randomiseringstoetsen in ver-



Figuur 2. Verloop depressiescores en interventiemoment (dag 15).



band met de effectiviteit van fitnessstraining om de depressieve gevoelens van Alex te beïnvloeden? In figuur 2 zien we hoe de 123 depressiescores door een interventie (fitheid) op dag 15 worden opgedeeld in een A- en een B-fase. We herhalen dat het interventiemoment in principe op grond van het toeval wordt bepaald.

Tabel 2. Geselecteerde output.

---

*Computerprogramma Single Case Randomization Tests; versie 1.1*

*Subject 1*

---

N: 123  
Mean : 1.325  
Std: 0.621  
Range: 1.000 – 4.000  
Autocorr: 0.476  
Slope: – 0.002  
Intercept: 1.450

Systematic generation of randomizations

104 Randomizations calculated

Statistic: [A-B]

PMean(A)-PMean(B)

Value of data: 0.520

3 Randomizations higher or equal

p-Value: 0.028846

Parameters for Condition A

N: 14  
Mean: 1.786  
Std: 0.579  
Range: 1.000 – 3.000  
Autocorr: 0.135  
Slope: 0.029  
Intercept: 1.571

Parameters for Condition B

N: 109  
Mean: 1.266  
Std: 0.603  
Range: 1.000 – 4.000  
Autocorr: 0.464  
Slope: 0.001  
Intercept: 1.217

---



Voor de 123 observaties en per fase krijgen we in tabel 2 enkele statistische gegevens. De gemiddelde depressiescore voor fase A is 1,786, voor fase B 1,266. Het verschil tussen deze gemiddelden is 0,52. Onder de nulhypothese dat de fitnessstraining geen effect heeft is de kans om een verschil van 0,52 te vinden gelijk aan 0,03 (afgerond). Deze kans is zo klein dat we de nulhypothese verworpen, en we besluiten dat fitnessstraining bij Alex in significante mate bijdraagt tot een vermindering van de depressie. Van de 104 mogelijke randomiseringën (alle mogelijke verschillen tussen A- en B-fase) zijn er slechts 3, waar het verkregen verschil gelijk is aan of groter is dan 0,52.

Fitnessstraining is voor Alex een efficiënte therapievorm om zijn depressieve gevoelens te reduceren. Het behoeft geen betoog dat het bespreken van deze bevindingen met de betrokken patiënt een grote motiverende waarde kan hebben. Fitnessstraining vergt een volgehouden inspanning. Het weten dat men goed bezig is betekent een extra stimulans. Het belangrijke is dat effectonderzoek niet louter een academische aangelegenheid is, maar ook therapeutisch relevant.

### **Tot slot**

Tijdreeksenanalyses en randomiseringstoetsen openen een aantal boeiende perspectieven bij de bestudering van één geval. In principe wordt therapeutisch maatwerk mogelijk. Wat is werkzaam bij deze patiënt? Er zijn sterke aanwijzingen dat fitnessstraining bij Alex effectief is in het reduceren van de depressieve gevoelens. Om therapeutische overwegingen kunnen de resultaten onder de vorm van feedback naar de patiënt worden teruggespeeld. Voor de patiënt kan dit zeer motiverend zijn. Dit past in een visie op de therapeutische relatie als een therapeutische alliantie, waarbij therapeut en patiënt, ieder vanuit hun eigen deskundigheid, samen op zoek zijn naar manieren om een probleem op doelmatige wijze te beïnvloeden. Dit veronderstelt kritische zin en bereidheid de uitgangspunten van zowel patiënt als therapeut kritisch te toetsen.

Naast de therapeutische relevantie is er ook de wetenschappelijke relevantie. We denken hier in eerste instantie aan outcome- en procesonderzoek van psychotherapie. De voorgestelde methodologie kan bijdragen tot het identificeren van therapeutisch werkzame factoren en tot het exploreren van wat er in een therapie gebeurt.

We zijn er ons van bewust dat de statistische methodologie, de theoretische achtergrond en de concrete toepassing te opper-



vlakkelig, onvolledig en wat simplifiërend weergegeven zijn. Het was onze bedoeling een illustratieve casus te brengen die de mogelijkheden van een onderzoeksmethode en de relevantie voor de klinische praktijk zou tonen. De fobische instelling van menig clinicus kennend voor al wat cijfers en tabellen betreft, hebben we afgezien van een conditionele stimulus in de vorm van een verantwoordwoord en grondiger situering, die maar zou leiden tot vermijdingsgedrag in de vorm van het voortijdig afhaken bij het lezen van dit artikel.

---

### Summary

The assessment of the effect of an intervention. A case illustration with application of time series analysis and randomization tests

The assessment of the outcome of psychotherapy is an important endeavor, with an impact on both clinical practice and research. The 'scientist-practitioner' model emphasizes the ability of therapists to apply scientific methods in the evaluation of their work. In an illustrative case, we study the impact of fitness training on depression. Data are analyzed by means of time series analysis and randomization tests. Both methods can be applied to the systematic study of individual cases and are particularly helpful to study outcome and process research. This approach is especially suited to behavior therapists accustomed to systematic inquiry.

**Key words:** case study, outcome research, time series analysis, randomization tests.

---

### Literatuur

- Buijs, A. (1994). Voorspellen met tijdreeksen. In *Statistiek om mee verder te werken* (pp. 199-251). Houten: Stenfert Kroese.
- De Boeck, P. (1986). Tijdreeksen-analyse in de context van therapie. *Gedrags therapie*, 19(1), 33-52.
- De Boeck, P. (1988). Het A-B-A-design met toepassing van tijdreeksenanalyse. In *Psychologie van het individu* (pp. 163-177). Assen/Maastricht: Van Gorcum/Universitaire Pers Leuven.
- De Boeck, P. (1988). Het A-B-design met toepassing van randomiseringstoetsen. In *Psychologie van het individu* (pp. 178-186). Assen/Maastricht: Universitaire Pers Leuven/Van Gorcum.
- Edgington, E.S. (1984). Statistics and single case analysis. In M. Hersen, R.M. Eisler & P.M. Miller (Eds.), *Progress in behavior modification* (pp. 16, 83-119). New York: Raven Press.



- Hilliard, R.B. (1993). Single-case methodology in psychotherapy process and outcome research. *Journal of Consulting and Clinical Psychology*, 61(3), 373-380.
- Kazdin, A.E. (1982). *Single-case research designs: Methods for clinical and applied settings*. Oxford/New York: Oxford University Press.
- Knapen, J. (1995). Fitnessbegeleiding van depressieve patiënten in een gedragstherapeutische setting. In J. Simons (red.), *Actuele themata uit de psychomotorische therapie* (pp. 37-45.). Leuven: Acco.
- Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative outcome studies of psychotherapy: is it true that 'everyone has won and all must have prizes?'. *Archives of General Psychiatry*, 32, 995-1008.
- Martinsen, E.W. (1995). The effects of exercise on mental health in clinical populations. In S.J.H. Biddle (Ed.), *European perspectives on exercise and sport psychology* (pp. 71-84). Champaign: Human Kinetics Publishers.
- McCleary, R., & Hay, R.A. (1980). *Applied time-series analysis for the social sciences*. Beverly Hills, CA: Sage.
- McLeod, J. (1994). *Doing counselling research*. Londen: Sage Publications.
- Onghena, P. (1992). Randomization tests for extensions and variations of ABAB single-case experimental designs: A rejoinder. *Behavioral Assessment*, 14(2), 153-171.
- Park, H.S., Marascuilo, L., & Gaylord, R.R. (1990). Visual inspection and statistical analysis in single-case designs. *Journal of Experimental Education*, 58(4), 311-320.
- SAS Institute (1993). *SAS/ETS user's guide, version 6* (2nd ed). Cary, NC: SAS Institute. 1022 pp.
- Sloane, R.B., Staples, F.R., Cristol, A.H., Yorkston, N.J., & Whipple, K. (1975). *Psychotherapy versus behavior therapy*. Cambridge, MA: Harvard University Press.
- Smith, M., Glass, G., & Miller, T. (1980). *The benefits of psychotherapy*. Baltimore, MD: John Hopkins University Press.
- Van Coppenolle, H., Pieters, G., Knapen, J., & Peuskens, J. (1993). Psychomotor therapy in depressive patients. *Issues in Special Education & Rehabilitation*, 8(2), 29-34.
- Van Damme, G., & Onghena, P. (1993). *Single case randomization tests, versie 1.1 (computerprogramma)*. Departement Psychologie, Katholieke Universiteit Leuven.
-